

# Statistical post-processing of forecasts for extremes using bivariate brown-resnick processes with an application to wind gusts

Marco Oesting<sup>1</sup> · Martin Schlather<sup>2</sup> ·  
Petra Friederichs<sup>3</sup>

Received: 1 September 2015 / Revised: 1 November 2016 / Accepted: 21 November 2016 /  
Published online: 13 December 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** To improve the forecasts of weather extremes, we propose a joint spatial model for the observations and the forecasts, based on a bivariate Brown-Resnick process. As the class of stationary bivariate Brown-Resnick processes is fully characterized by the class of pseudo cross-variograms, we contribute to the theoretical understanding of pseudo cross-variograms refining the knowledge of the asymptotic behaviour of all their components and introducing a parsimonious, but flexible parametric model. Both findings are of interest in classical geostatistics on their own. The proposed model is applied to real observation and forecast data for extreme wind gusts at 119 stations in Northern Germany.

**Keywords** Bivariate random field · Matérn model · Max-stable process · Pseudo cross-variogram

**AMS 2000 Subject Classifications** 60G70 · 62M30 · 60G60

## 1 Introduction

Spatial extremes may occur in various forms such as heavy rainfall, floods, heat waves or wind gusts. In view of their severe consequences, an adequate and precise

---

✉ Marco Oesting  
oesting@mathematik.uni-siegen.de

<sup>1</sup> Department Mathematik, Universität Siegen, Walter-Flex-Str. 3, D-57072 Siegen, Germany

<sup>2</sup> Institut für Mathematik, Universität Mannheim, A5, 6, D-68131 Mannheim, Germany

<sup>3</sup> Meteorologisches Institut, Universität Bonn, Auf dem Hügel 20, D-53121 Bonn, Germany

forecast of these events is of great importance. However, the rareness of extreme events impedes any such task and, consequently, existing forecasts often lack accuracy. In meteorology, for example, forecasting extreme wind gusts, which are defined as peak wind speeds over a few seconds, is exacerbated by the short temporal and spatial ranges. Furthermore, numerical weather prediction (NWP) models provide estimates or diagnoses of wind gusts based on empirical knowledge only (cf. Brasseur 2001). Although wind is a prognostic variable in NWP models, its value represents an average wind speed over a few minutes or longer depending on the grid spacing of the NWP model. Hence, post-processing procedures are needed that allow for an enhanced probabilistic forecast.

Occurring as limits of normalized pointwise maxima of stochastic processes, max-stable processes provide a suitable framework for the description of spatial extreme events, commonly used in environmental sciences (Coles 1993; Coles and Tawn 1996; Huser and Davison 2014). Of particular interest is the subclass formed by Brown-Resnick processes which arise as limits of rescaled maxima of Gaussian processes (Brown and Resnick 1977; Kabluchko et al. 2009; Kabluchko 2011).

During the last years, max-stable processes have been frequently applied as models for spatial extremes in environmental sciences. For instance, Engelke et al. (2015) and Genton et al. (2015) recently used max-stable processes to model extreme wind speed observations. The model we propose will go one step further, also taking into account the forecasts in two different aspects: First and in contrast to Engelke et al. (2015) and Genton et al. (2015), we consider the mean forecast to get a normalized version of the extreme observations. Second, besides the observable variable of interest itself, the corresponding forecast is included as second variable yielding a bivariate max-stable process. Here, we will focus on the class of bivariate Brown-Resnick processes (cf. Molchanov and Stucki 2013; Genton et al. 2015) to exploit the statistical relation between observable data and the corresponding forecast. Modeling the behavior of observational data, a sample from the distribution of the observations conditional on the forecast is supposed to provide more realistic results than the original forecast and thus will appear as an appropriate probabilistic post-processed forecast.

The paper is structured as follows: In Section 2, we present a univariate model for extreme observations, which may, in general, provide a first alternative to the original forecast. We introduce a model for the marginal distribution, i.e. the distribution of the observable variable of interest at a single location, motivating the normalization of its extremes by the mean forecast. The spatial dependence structure is incorporated into the model by the use of univariate Brown-Resnick processes. Section 3 is dedicated to the bivariate Brown-Resnick process which serves as a joint model for both the maximally observed and forecasted quantities. We deduce a necessary condition on the asymptotic behavior of the pseudo cross-variogram and provide a flexible cross-variogram model which leads to a stationary bivariate Brown-Resnick process. In Section 4, we describe how the model can be fitted to data. Based on this model, we propose a post-processing procedure which is presented in Section 5. Further, we provide tools to verify the procedure and the underlying models. Finally,

the methods presented in Sections 4 and 5 are applied to real observation and forecast data for extreme wind gusts provided by Germany's National Meteorological Service, Deutscher Wetterdienst (DWD) (Section 6).

## 2 Modeling by a univariate random field

In this section, we present a spatial model for the observed pointwise maximum  $V_{\max}^{\text{obs}}$  within a specific time period. To this end, we assume that, for each location and time period, the maximum  $V_{\max}^{\text{obs}}$  is based on observations at  $N$  equidistant instants of times per period, that is, we have  $V_{\max}^{\text{obs}} = \max_{t=1, \dots, N} V_t^{\text{obs}}$  for  $V_1^{\text{obs}}, \dots, V_N^{\text{obs}} \sim F_{\vartheta}$  for some parameter  $\vartheta$ . Here, the probability distributions  $F_{\vartheta}$  are supposed to form a location-scale family with finite second moments, i.e.  $\vartheta = (m, s) \in \mathbb{R} \times (0, \infty)$  with  $F_{(m,s)}(x) = F_{(0,1)}\left(\frac{x-m}{s}\right)$ ,  $x \in \mathbb{R}$ , and  $F_{(0,1)}$  is standardized to mean zero and unit variance. We assume  $\vartheta = (m, s)$  to be temporally constant at each location within the same time period, but allow the values to vary among different locations and different time periods. The values of  $m$  and  $s$  will essentially be estimated from the bulk of the distribution, not the tail, and thus, they can often be extracted accurately from forecasts. Within the same time period and at the same location, the observable variables  $V_1^{\text{obs}}, \dots, V_N^{\text{obs}}$  are assumed to be subsequent  $N$  elements of a stationary time series  $(V_t^{\text{obs}})_{t \in \mathbb{Z}}$ . Furthermore, we assume that the standardized distribution  $F_{(0,1)}$  belongs to the max-domain of attraction of some univariate extreme value distribution  $G_{\xi}$ ,  $\xi \in \mathbb{R}$ , i.e. there are sequences  $(a_n)_{n \in \mathbb{N}}$ ,  $a_n > 0$ , and  $(b_n)_{n \in \mathbb{N}}$ ,  $b_n \in \mathbb{R}$ , such that

$$F_{(0,1)}^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G_{\xi}(x), \quad 1 + \xi x > 0,$$

where

$$G_{\xi}(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}), & \xi \neq 0, \\ \exp(-\exp(-x)), & \xi = 0, \end{cases}$$

for  $1 + \xi x > 0$ . As the second moment of  $F_{(0,1)}$  is assumed to be finite, we have  $\xi < 0.5$ . Under some conditions on the regularity and the dependence of the stationary sequence  $V_1^{\text{obs}}, V_2^{\text{obs}}, \dots$ , we obtain that

$$\mathbb{P}\left(\frac{\max_{i=1, \dots, n} V_i^{\text{obs}} - m - \tilde{b}_n s}{\tilde{a}_n s} \leq x\right) \xrightarrow{n \rightarrow \infty} G_{\xi}(x), \quad 1 + \xi x > 0, \quad (1)$$

where  $\tilde{a}_n = a_n \theta^{-\xi}$  and  $\tilde{b}_n = b_n - \xi^{-1}(1 - \theta^{-\xi})$  for some  $\theta \in (0, 1]$  called extremal index (cf. Coles 2001; Leadbetter et al 1983).

Let  $m = m(l, p)$  and  $s = s(l, p)$  be the mean of the variable at location  $l$  and period  $p$  and its standard deviation, respectively. Let

$$G_{\xi, \mu, \sigma}(x) = G_{\xi}((x - \mu)/\sigma), \quad 1 + \xi(x - \mu)/\sigma > 0$$

be the generalized extreme value distribution (GEV). Then, considering the maximum  $V_{\max}^{\text{obs}} = V_{\max}^{\text{obs}}(l, p)$  for large  $N$ , we have approximately that

$$\frac{V_{\max}^{\text{obs}}(l, p) - m(l, p)}{s(l, p)} \sim G_{\xi^{\text{obs}}, \mu^{\text{obs}}(l), \sigma^{\text{obs}}(l)}. \quad (2)$$

Here, the GEV parameters are assumed to be the same for every time period, which, in general, enables us to estimate the parameters for current and future time periods from past data. As common in many applications, the extreme value index  $\xi$  is also assumed to be constant in space. Under the ideal assumption that  $V_i^{\text{obs}} \sim F_{(m,s)}$  and that  $m$  and  $s$  can be determined exactly, the GEV parameters  $\xi^{\text{obs}}$ ,  $\mu^{\text{obs}}$  and  $\sigma^{\text{obs}}$  are constant in space, as well. In practice, however, the observed variable of interest is subject to measurement errors whose distribution is spatially varying. Further,  $m$  and  $s$  often need to be extracted from forecasts with limited spatial resolution. To account for these difficulties, we allow  $\mu^{\text{obs}}(l)$  and  $\sigma^{\text{obs}}(l)$  to depend on the location  $l$ , while the extreme value index  $\xi^{\text{obs}}$  is assumed to be constant in space, as common in many applications. In contrast to  $\mu^{\text{obs}}$  and  $\sigma^{\text{obs}}$ ,  $m(l, p)$  and  $s(l, p)$  vary in space and time and may be interpreted as normalizing constants that will be the same for observation and forecasts. As  $m(l, p)$  and  $s(l, p)$  are defined as mean and standard deviation of the variable of interest, the parameters  $\mu^{\text{obs}}(l)$  and  $\sigma^{\text{obs}}(l)$  are uniquely determined. Marginal transformation yields that

$$X^{\text{obs}}(l, p) = \frac{1}{\xi^{\text{obs}}} \log \left( 1 + \xi^{\text{obs}} \frac{V_{\max}^{\text{obs}}(l, p) - m(l, p) - s(l, p)\mu^{\text{obs}}(l)}{s(l, p)\sigma^{\text{obs}}(l)} \right) \quad (3)$$

is standard Gumbel distributed for every location  $l$  and time period  $p$ .

Perceiving the set of locations as a subset of  $\mathbb{R}^2$  and the set of periods as a subset of  $\mathbb{Z}$ , the transformed observations can be regarded as realizations of a spatio-temporal random field  $\{X^{\text{obs}}(l, p), l \in \mathbb{R}^2, p \in \mathbb{Z}\}$ . While we assume that the spatial random fields  $\{X^{\text{obs}}(l, p), l \in \mathbb{R}^2\}$ ,  $p \in \mathbb{Z}$ , are independent and identically distributed, we allow for a non-trivial spatial dependence structure. Here, we use the class of Brown-Resnick processes that can be defined for arbitrary dimensions  $D$  (Brown and Resnick 1977; Kabluchko et al. 2009): Let  $\Pi = \sum_{i \in \mathbb{N}} \delta_{U_i}$  be a Poisson point process on  $\mathbb{R}$  with intensity  $e^{-u} du$  and, independently of  $\Pi$ , let  $W_i$ ,  $i \in \mathbb{N}$ , be independent copies of a zero-mean Gaussian random field  $\{W(s), s \in \mathbb{R}^D\}$  with stationary increments and semi-variogram  $\gamma(\cdot)$  defined by

$$2\gamma(s) = \text{Var}(W(s) - W(0)), \quad s \in \mathbb{R}^D.$$

Then, the random field  $Z$  defined by

$$Z(s) = \max_{i \in \mathbb{N}} (U_i + W_i(s) - \text{Var}(W(s))/2), \quad s \in \mathbb{R}^D,$$

and called Brown-Resnick process associated to the semi-variogram  $\gamma$ , is stationary and max-stable with standard Gumbel margins and its law only depends on the semi-variogram  $\gamma$  (Kabluchko et al. 2009). For the application of the Brown-Resnick model to observed data with locations in  $\mathbb{R}^2$ , we propose to restrict to semi-variograms from a flexible parametric subclass, such as semi-variograms of the type

$$\gamma_{\vartheta}(h) = \|sA(b, \zeta)h\|^{\alpha}, \quad h \in \mathbb{R}^2, \quad (4)$$

with  $\vartheta = (s, b, \zeta, \alpha)$  for  $s, b > 0$ ,  $\zeta \in (-\pi/4, \pi/4]$  and  $\alpha \in (0, 2]$ . Here, the matrix  $A(b, \zeta) \in \mathbb{R}^{2 \times 2}$  allows for geometric (elliptical) anisotropy, i.e.

$$A(b, \zeta) = \begin{pmatrix} \cos \zeta & \sin \zeta \\ -b \sin \zeta & b \cos \zeta \end{pmatrix} \quad (5)$$

(? [, cf.)Subsection 2.5.2]chiles-delfiner-2012, and  $s$  is an overall scale factor.

### 3 Modeling by a bivariate random field

In this section, we also take into account the dependence between the observed maximum  $V_{\max}^{\text{obs}}$  and its forecast  $V_{\max}^{\text{pred}}$ . As  $V_{\max}^{\text{pred}}$  is a forecast for  $V_{\max}^{\text{obs}}$ , it seems reasonable to use a GEV model similar to the one described in Section 2 with possibly different parameters  $\xi^{\text{pred}}$ ,  $\mu^{\text{pred}}(\cdot)$  and  $\sigma^{\text{pred}}(\cdot)$ , i.e.

$$\frac{V_{\max}^{\text{pred}}(l, p) - m(l, p)}{s(l, p)} \sim G_{\xi^{\text{pred}}, \mu^{\text{pred}}(l), \sigma^{\text{pred}}(l)} \quad (6)$$

(cf. Eq. 2). Marginally transforming  $V_{\max}^{\text{pred}}$  analogously to Eq. 3 yields a random field  $\{X^{\text{pred}}(l, p), l \in \mathbb{R}^2, p \in \mathbb{Z}\}$  with standard Gumbel margins. Thus, we end up with bivariate spatial random fields  $\{(X^{\text{obs}}(l, p), X^{\text{pred}}(l, p)), l \in \mathbb{R}^2\}$  which are assumed to be independent and identically distributed for  $p \in \mathbb{Z}$ .

A bivariate Brown-Resnick process can be constructed in the following way (cf. Molchanov and Stucki 2013; Genton et al 2015): Let  $\sum_{i \in \mathbb{N}} \delta_{U_i}$  be a Poisson point process on  $\mathbb{R}$  with intensity measure  $e^{-u} du$ . Further, let  $W_i$ ,  $i \in \mathbb{N}$ , be independent copies of a bivariate centered Gaussian process  $W = (W^{(1)}, W^{(2)})^\top = \{(W^{(1)}(s), W^{(2)}(s))^\top : s \in \mathbb{R}^D\}$  such that the pseudo cross-variogram (Clark et al. 1989; Papritz et al. 1993)  $\gamma(h) = (\gamma_{ij}(h))_{i,j \in \{1,2\}}$  defined by

$$2\gamma_{ij}(h) = \text{Var}(W^{(i)}(s+h) - W^{(j)}(s)), \quad h \in \mathbb{R}^D, \quad (7)$$

does not depend on  $s \in \mathbb{R}^D$ . Analogously to the univariate Brown-Resnick process, it can be shown that the bivariate Brown-Resnick process  $Z = (Z^{(1)}, Z^{(2)})^\top$  defined by

$$Z^{(j)}(s) = \max_{i \in \mathbb{N}} (U_i + W_i^{(j)}(s) - \text{Var}(W^{(j)}(s))/2), \quad s \in \mathbb{R}^D, \quad j = 1, 2, \quad (8)$$

is max-stable and stationary. Its law only depends on the pseudo cross-variogram  $\gamma$ .

*Remark 1* The fact that  $(\gamma_{ij}(h))_{i,j=1,2}$  can be defined independently of  $s \in \mathbb{R}^D$  implies that  $W$  is intrinsically stationary, i.e. the process  $\{W(s+h) - W(s) : s \in \mathbb{R}^D\}$  is stationary for every  $h \in \mathbb{R}^D$ . Both conditions, however, are not equivalent as the definition of  $\gamma_{12}(h)$  might depend on  $s \in \mathbb{R}^D$  even if  $W$  is intrinsically stationary. For instance, if both components are independent, we have  $\gamma_{12}(h) = 2\gamma_{11}(s+h) + 2\gamma_{22}(s)$  for the off-diagonal element of pseudo cross-variogram. By way of contrast, intrinsic stationarity is equivalent to the cross variogram  $h \mapsto (\mathbb{E}(W_i(s+h) - W_i(s))(W_j(s+h) - W_j(s)))_{i,j=1,2}$  being independent of  $s$ .

Indeed, Molchanov and Stucki (2013) already gave necessary and sufficient conditions for a multivariate process of Brown-Resnick type to be stationary. For a fixed intensity  $e^{-u} du$  of the Poisson point process, the conditions on Gaussian processes given in Theorem 5.3 in Molchanov and Stucki (2013) can be shown to be equivalent to the conditions on the process  $W$  stated above (if we additionally require  $Z$  to have standard Gumbel margins) by a straightforward computation. Thus, the Gaussian processes in the above definition of bivariate Brown-Resnick processes are essentially the only ones that yield a stationary max-stable process.

In the following, we investigate the structure and the asymptotic behavior of bivariate variograms that are translation invariant, refining the result by Papritz et al. (1993) that  $\lim_{h \rightarrow \infty} \gamma_{12}(h)/\gamma_{11}(h) = 1$  if  $\gamma_{11}$  is unbounded. This allows us to find valid models for bivariate Brown-Resnick processes. The following theorem, as well as the statements above, immediately extend to the general multivariate case.

**Theorem 1** *Let  $W = (W^{(1)}, W^{(2)})^\top$  be a bivariate second-order process on  $\mathbb{R}^D$  with pseudo cross-variogram  $(\gamma_{ij}(h))_{i,j \in \{1,2\}}$  which does not depend on  $s \in \mathbb{R}^D$ . Then,*

$$\sqrt{\gamma(h)} = \sqrt{(\gamma_{ij}(h))_{i,j \in \{1,2\}}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \sqrt{\gamma_0(h)} + \begin{pmatrix} f_{11}(h) & f_{12}(h) \\ f_{21}(h) & f_{22}(h) \end{pmatrix}$$

for some univariate variogram  $\gamma_0$  and bounded functions  $f_{11}, f_{12}, f_{21}, f_{22} : \mathbb{R}^D \rightarrow \mathbb{R}$ .

*Proof* For  $i, j \in \{1, 2\}$ , and  $h \in \mathbb{R}^D$ , we obtain

$$\begin{aligned} \left( \sqrt{\gamma_{ii}(h)} - \sqrt{\gamma_{jj}(h)} \right)^2 &= \gamma_{ii}(h) - 2\sqrt{\gamma_{ii}(h)\gamma_{jj}(h)} + \gamma_{jj}(h) \\ &\leq \gamma_{ii}(h) - \text{Cov}(W^{(i)}(h) - W^{(i)}(0), W^{(j)}(h) - W^{(j)}(0)) + \gamma_{jj}(h) \\ &= \frac{1}{2} \text{Var}(W^{(i)}(h) - W^{(i)}(0) - W^{(j)}(h) + W^{(j)}(0)) \\ &= \gamma_{ij}(0) - \text{Cov}(W^{(i)}(h) - W^{(j)}(h), W^{(i)}(0) - W^{(j)}(0)) + \gamma_{ij}(0) \leq 4\gamma_{ij}(0), \end{aligned}$$

where we used the Cauchy-Schwarz inequality for both inequalities. Analogously, we get the assessment

$$\begin{aligned} \left( \sqrt{\gamma_{ii}(h)} - \sqrt{\gamma_{ji}(h)} \right)^2 &= \gamma_{ii}(h) - 2\sqrt{\gamma_{ii}(h)\gamma_{ji}(h)} + \gamma_{ji}(h) \\ &\leq \gamma_{ii}(h) - \text{Cov}(W^{(i)}(h) - W^{(i)}(0), W^{(j)}(h) - W^{(i)}(0)) + \gamma_{ji}(h) \\ &= \frac{1}{2} \text{Var}(W^{(i)}(h) - W^{(j)}(h)) = \gamma_{ij}(0). \end{aligned}$$

Thus, the assertion of the theorem follows with  $\gamma_0 = \gamma_{11}$ .  $\square$

As the components of a translation invariant bivariate pseudo cross-variogram only differ by a function that may increase only with a rate of order  $O(\sqrt{\gamma_0(h)})$  (Theorem 1), a reasonable and not too restrictive model for the corresponding bivariate Gaussian random field  $W = (W^{(1)}, W^{(2)})^\top$  is given by

$$W(s) = (1, 1)^\top V_1(s) + V_2(s), \quad s \in \mathbb{R}^D,$$

where  $V_1$  is a univariate Gaussian random field with stationary increments and semi-variogram  $\gamma_0$  and  $V_2$  is a bivariate stationary Gaussian random field with bivariate cross-covariance function  $C(h) = (C_{ij}(h))_{i,j \in \{1,2\}}$ , independent from  $V_1$ . Then, the pseudo cross-variogram  $\gamma$  of  $W$  has the form

$$\gamma_{ij}(h) = \gamma_0(h) + \frac{1}{2}C_{ii}(0) + \frac{1}{2}C_{jj}(0) - C_{ij}(h), \quad i, j \in \{1, 2\}, \quad h \in \mathbb{R}^D.$$

Analogously to the univariate case, we propose to restrict to a parametric subclass of semi-variograms for  $\gamma_0$  such as

$$\gamma_0(h) = \sigma^2 \frac{(\kappa^{-1}\|h\|)^2}{((\kappa^{-1}\|h\|)^2 + 1)^\beta}$$

where  $\sigma, \kappa > 0$  and  $\beta \in (0, 1)$ . Here,  $\gamma_0$  is a valid univariate variogram as  $h \rightarrow \|h\|^2$  is a variogram and  $\lambda \mapsto \lambda/(\lambda + 1)^\beta$  is a Bernstein function (cf. Berg et al 1984; Schilling et al 2010). Note that  $\gamma_0$  is a variogram of power law type modified to be smooth at the origin.

For the bivariate cross-covariance  $C$ , we propose to use a parsimonious version of the bivariate Matérn model (cf. Gneiting et al 2010), which is a bivariate generalization of one of the most widely used models in geostatistics, the Matérn model (cf. Guttorp and Gneiting 2006; Stein 1999, for example). In the bivariate Matérn model, each component of  $C$  is a Matérn covariance function which we parametrize in the way suggested by Handcock and Wallis (1994), i.e.

$$\begin{aligned} C_{ii}(h) &= \sigma_i^2 \frac{2^{1-v_i}}{\Gamma(v_i)} \left( \frac{2\sqrt{v_i}}{a_i} \|h\| \right)^{v_i} K_{v_i} \left( \frac{2\sqrt{v_i}}{a_i} \|h\| \right), \quad i = 1, 2, \\ C_{12}(h) = C_{21}(h) &= \rho \sigma_1 \sigma_2 \frac{2^{1-v_{12}}}{\Gamma(v_{12})} \left( \frac{2\sqrt{v_{12}}}{a_{12}} \|h\| \right)^{v_{12}} K_{v_{12}} \left( \frac{2\sqrt{v_{12}}}{a_{12}} \|h\| \right), \end{aligned}$$

for  $a_1, a_2, a_{12}, \sigma_1, \sigma_2, v_1, v_2, v_{12} > 0$  and suitable  $\rho \in [-1, 1]$ .

Here, analogously to the parsimonious version of the bivariate Matérn model which is based on a different parametrization (Gneiting et al. 2010), we set  $a_1 = a_{12} = a_2 = a > 0$  and  $v_{12} = \frac{1}{2}(v_1 + v_2)$ . Then, by Thm. 3 in Gneiting et al. (2010),  $C$  is a valid bivariate cross-covariance model if and only if

$$\rho^2 \leq \frac{(1 + v_{12}^{-1})^{2v_{12}+2}}{(1 + v_1^{-1})^{v_1+1}(1 + v_2^{-1})^{v_2+1}}.$$

To increase the flexibility of the model, we further add a spatially constant effect with variance  $c^2$  in the second component. Thus,  $C$  has the form

$$\begin{aligned} C_{11}(h) &= \sigma_1^2 \frac{2^{1-v_1}}{\Gamma(v_1)} \left( \frac{2\sqrt{v_1}}{a} \|h\| \right)^{v_1} K_{v_1} \left( \frac{2\sqrt{v_1}}{a} \|h\| \right), \\ C_{12}(h) = C_{21}(h) &= \rho \sigma_1 \sigma_2 \frac{2^{1-v_{12}}}{\Gamma(v_{12})} \left( \frac{2\sqrt{v_{12}}}{a} \|h\| \right)^{v_{12}} K_{v_{12}} \left( \frac{2\sqrt{v_{12}}}{a} \|h\| \right), \\ C_{22}(h) &= c^2 + \sigma_2^2 \frac{2^{1-v_2}}{\Gamma(v_2)} \left( \frac{2\sqrt{v_2}}{a} \|h\| \right)^{v_2} K_{v_2} \left( \frac{2\sqrt{v_2}}{a} \|h\| \right). \end{aligned}$$

Note that as the common summand  $\gamma_0$  is smooth at the origin, the behavior of  $\gamma_{ii}$  near the origin, i.e. the differentiability of  $W^{(i)}$ , depends only on the behavior of  $C$  which can be modeled flexibly by the smoothness parameters  $\nu_1$  and  $\nu_2$  of the bivariate Matérn model. In particular, as  $\|h\| \rightarrow 0$  and for some  $k(a, \nu) > 0$ , we have

$$\gamma_{ii}(h) = \begin{cases} k(a, \nu_i) \|h\|^{2\nu_i} + O(\|h\|^2), & \nu_i < 1, \\ k(a, 1) \|h\|^2 \log \|h\| + O(\|h\|^2), & \nu_i = 1, \\ k(a, \nu_i) \|h\|^2 + o(\|h\|^2), & \nu_i > 1 \end{cases}$$

(cf. Stein 1999). Furthermore, the sample paths are  $m$  times differentiable if and only if  $\nu > m$  (Gelfand et al. 2010). The behavior of the  $\gamma_{ii}$  as  $\|h\| \rightarrow \infty$ , which has to be the same for all components by Theorem 1, is parameterized by  $\beta$  as we have  $\gamma_{ii}(h) \|h\|^{-2(1-\beta)} \rightarrow 1$  as  $\|h\| \rightarrow \infty$ . To increase the applicability of our model to real data located in  $\mathbb{R}^2$ , we further allow for geometric anisotropy, replacing  $\|h\|$  by  $\|h^*\|$  where  $h^* = A(b, \zeta)h$  and  $A(b, \zeta)$  is the anisotropy matrix defined in Eq. 5. Thus, we obtain the variogram model  $\gamma(\vartheta; \cdot)$  given by

$$\begin{aligned} \gamma_{ii}(\vartheta; h) &= \sigma^2 \frac{(\kappa^{-1} \|h^*\|)^2}{((\kappa \|h^*\|)^2 + 1)^\beta} + \sigma_i^2 \left( 1 - \left( \frac{2\sqrt{\nu_i}}{a} \|h^*\| \right)^{\nu_i} K_{\nu_i} \left( \frac{2\sqrt{\nu_i}}{a} \|h^*\| \right) \right), \\ \gamma_{12}(\vartheta; h) &= \sigma^2 \frac{(\kappa^{-1} \|h^*\|)^2}{((\kappa \|h^*\|)^2 + 1)^\beta} + \frac{\sigma_1^2 + c^2 + \sigma_2^2}{2} \\ &\quad - \rho \sigma_1 \sigma_2 \frac{2^{1-\nu_{12}}}{\Gamma(\nu_{12})} \left( \frac{2\sqrt{\nu_{12}}}{a} \|h^*\| \right)^{\nu_{12}} K_{\nu_{12}} \left( \frac{2\sqrt{\nu_{12}}}{a} \|h^*\| \right), \end{aligned} \quad (9)$$

for  $i = 1, 2$  and  $h \in \mathbb{R}^2$  where  $\vartheta = (\sigma, \kappa, b, \zeta, \beta, c, \sigma_1, \nu_1, \sigma_2, \nu_2, a, \rho)$ .

## 4 Model fitting

In the following, we will assume that data  $v_{\max}^{\text{obs}}(l_i, p)$  and  $v_{\max}^{\text{pred}}(l_i, p)$  for the maximal observed and forecasted variable of interest at stations  $l_i$ ,  $i = 1, \dots, n_l$ , and time period  $p = 1, \dots, n_p$  are available.

### 4.1 Fitting of the univariate model

Let henceforth be  $k \in \{\text{"obs"}, \text{"pred"}\}$ . We concentrate here on the estimation of the GEV and max-stable parameters assuming that the unknown mean  $m(l_i, d)$  and standard deviation  $s(l_i, p)$  of the underlying distribution  $F$  have already been estimated by  $\hat{m}(l_i, p)$  and  $\hat{s}(l_i, p)$ , respectively. An example for the later estimates can be found in Section 6. Given the estimates  $\hat{m}(l_i, p)$  and  $\hat{s}(l_i, p)$ , we obtain the standardized data

$$y^k(l_i, p) = \frac{v_{\max}^k(l_i, p) - \hat{m}(l_i, p)}{\hat{s}(l_i, p)}, \quad i = 1, \dots, n_l, \quad p = 1, \dots, n_p, \quad (10)$$

which are assumed to be GEV distributed with parameters  $\xi^k$ ,  $\mu^k(l_i)$  and  $\sigma^k(l_i)$ . We assume that the parameters are independent between the stations. This can be justified by measurement errors and the fact, that the forecasts used to estimate



$\hat{m}(l_i, d)$  and  $\hat{s}(l_i, d)$  are not directly available for the station  $l_i$  but only for the closest grid point. Furthermore, we face model errors, e.g. misrepresentation of orographic effects. Effects stemming from the environment of the measurement stations might even be the major cause for the variations. As genuine variation and measurement errors cannot be separated in our set up, we estimate the parameters separately for each station, via maximum likelihood. As the standardized data  $y^k$  are assumed to be temporally independent, by Smith (1985), the maximum likelihood estimators  $(\hat{\xi}^k(l_i), 1 \leq i \leq n_l)$  are asymptotically normally distributed if  $\xi^k > -0.5$ . Thus, under the hypothesis that  $\hat{\xi}^k = \frac{1}{n_l} \sum_{i=1}^{n_l} \hat{\xi}^k(l_i)$  is the true shape parameter of the GEV at each station, the standardized residuals

$$\frac{\hat{\xi}^k(l_1) - \hat{\xi}^k}{(\widehat{\text{Var}}(\hat{\xi}^k(l_1)))^{1/2}}, \dots, \frac{\hat{\xi}^k(l_{n_l}) - \hat{\xi}^k}{(\widehat{\text{Var}}(\hat{\xi}^k(l_{n_l})))^{1/2}}$$

are approximately standard normally distributed, where  $\widehat{\text{Var}}(\hat{\xi}^k(l_i))$  is the variance of  $\hat{\xi}^k(l_i)$  estimated via the Hesse matrix of the log-likelihood function. Thus, the three hypotheses that the shape parameter, the location and the scale parameter are spatially constant can be checked indirectly via one-sample Kolmogorov-Smirnov tests of the corresponding residuals for the standard normal distribution. Here, although the data for different locations may be dependent, we assume that the normalized estimated parameters are independent.

By transformation (3), the estimates  $\hat{\xi}^k, \hat{\mu}^k(l_i)$  and  $\hat{\sigma}^k(l_i)$  yield normalized data

$$x^k(l_i, p) = \frac{1}{\hat{\xi}^k} \log \left( 1 + \hat{\xi}^k \frac{y^k(l_i, p) - \hat{\mu}^k(l_i)}{\hat{\sigma}^k(l_i)} \right), \quad 1 \leq i \leq n_l, \quad 1 \leq p \leq n_p. \quad (11)$$

These can be compared to a standard Gumbel distribution via Kolmogorov-Smirnov tests separately for each station as a goodness-of-fit test for the marginal model

$$V_{\max}^k(l_i, p) \sim G_{\hat{\xi}^k, \hat{\mu}_v^k(l_i, p), \hat{\sigma}_v^k(l_i, p)}, \quad (12)$$

$$\text{where } \hat{\mu}_v^k(l, p) = \hat{m}(l, p) + \hat{s}(l, p)\hat{\mu}^k(l) \text{ and } \hat{\sigma}_v^k(l, p) = \hat{s}(l, p)\hat{\sigma}^k(l). \quad (13)$$

In order to capture the spatial dependence structure, a univariate Brown-Resnick process associated to a variogram  $\gamma^k$  as defined in Eq. 4 is fitted to the transformed data  $(x^k(l_i, p))_{1 \leq i \leq n_l, 1 \leq p \leq n_p}$ . Note that there exist numerous methods of inference for Brown-Resnick processes, see, for example, Engelke et al. (2015) for a comparison of different estimators. The method we will use is based on the extremal coefficient function (Schlather and Tawn 2003). For a stationary Brown-Resnick process associated to the semi-variogram  $\gamma^k$ , the pairwise extremal coefficients are given by

$$\begin{aligned} \theta^k(s_1, s_2) &= \frac{\log \mathbb{P}(X^k(s_1) \leq x, X^k(s_2) \leq x)}{\log \mathbb{P}(X^k(s_1) \leq x)} \\ &= 2\Phi \left( \sqrt{\gamma^k(s_1 - s_2)/2} \right), \quad s_1, s_2 \in \mathbb{R}^2, \end{aligned} \quad (14)$$

where  $\Phi$  denotes the standard normal distribution function (cf. Kabluchko et al. 2009). This relation can be used for fitting Brown-Resnick processes to real data as the extremal coefficients  $\theta^k(s_1, s_2)$  can be estimated well via the relation

$$\theta^k(s_1, s_2) = \frac{1 + 2\nu^{F,k}(s_1, s_2)}{1 - 2\nu^{F,k}(s_1, s_2)}, \quad s_1, s_2 \in \mathbb{R}^2, \quad (15)$$

where the  $F$ -madogram  $\nu^{F,k}(s_1, s_2)$  is defined by

$$\nu^{F,k}(s_1, s_2) = \frac{1}{2} \mathbb{E} \left| F(X^k(s_1)) - F(X^k(s_2)) \right|, \quad s_1, s_2 \in \mathbb{R}^2, \quad (16)$$

and  $F$  is the marginal distribution function of  $X^k(s)$  (Cooley et al. 2006). Thus, we obtain a plug-in estimator  $\hat{\theta}^k(l_i, l_j)$  for the extremal coefficients  $\theta^k(l_i, l_j)$ , by replacing  $\nu^{F,k}$  in Eq. 15 by an estimator  $\hat{\nu}^{F,k}(l_i, l_j)$ ,  $1 \leq i, j \leq n_l$ . In order to avoid propagation of errors in marginal modeling, we choose the non-parametric estimator

$$\hat{\nu}^{F,k}(l_i, l_j) = \frac{1}{2 \cdot n_p \cdot (n_p - 1)} \sum_{p=1}^{n_p} \left| R_p(x^k(l_i, \cdot)) - R_p(x^k(l_j, \cdot)) \right| \quad (17)$$

where  $R_p(x)$  denotes the rank of the  $p$ -th component of some vector  $x$  (cf. Ribatet 2013). Then, the corresponding variogram parameter vector  $\hat{\vartheta}^k$  can be estimated by a weighted least squares fit of  $\hat{\theta}^k(l_i, l_j)$  to  $\theta^k(l_i, l_j)$  as given in Eq. 14. As proposed by Smith (1990), we choose weights that depend on the (estimated) variance  $\widehat{\text{Var}}(\theta^k(l_i, l_j))$  of the estimator  $\theta^k(l_i, l_j)$ . Thus, we obtain the estimator

$$\hat{\vartheta}^k = \arg \min_{\vartheta} \sum_{1 \leq i < j \leq n_l} \left( \frac{\hat{\theta}^k(l_i, l_j) - 2\Phi\left(\sqrt{\gamma^k(l_i - l_j)/2}\right)}{\sqrt{\widehat{\text{Var}}(\theta^k(l_i, l_j))}} \right)^2. \quad (18)$$

We will further discuss the estimation of the variance of  $\theta^k(l_i, l_j)$  in Section 6.

## 4.2 Fitting of the bivariate model

For fitting the bivariate Brown-Resnick process  $\{(X^{\text{obs}}(l), X^{\text{pred}}(l))^{\top} : l \in \mathbb{R}^2\}$  we consider the extremal coefficients  $\theta^{k_1, k_2}(s, t)$  of max-stable vectors  $(X^{k_1}(s), X^{k_2}(t))^{\top}$  for  $k_1, k_2 \in \{\text{"obs"}, \text{"pred"}\}$ . The extremal coefficients can be estimated from the transformed data  $x^{\text{obs}}(l_i, p)$  and  $x^{\text{pred}}(l_i, p)$ ,  $1 \leq i \leq n_l$ ,  $1 \leq p \leq n_p$ , in the same way as in the univariate case. The resulting estimates  $\hat{\theta}^{k_1, k_2}(l_i, l_j)$ ,  $1 \leq i, j \leq n_l$ ,  $k_1, k_2 \in \{\text{"obs"}, \text{"pred"}\}$  are compared to the corresponding extremal coefficients of a bivariate Brown-Resnick process associated to the variogram  $\gamma(\vartheta; \cdot)$  yielding the weighted least squares fit

$$\hat{\vartheta} = \arg \min_{\vartheta} \sum_{1 \leq i, j \leq n_l} \sum_{k_1, k_2 \in \{\text{"obs"}, \text{"pred"}\}} \left( \frac{\hat{\theta}^{k_1, k_2}(l_i, l_j) - 2\Phi\left(\sqrt{\gamma_{k_1, k_2}(\vartheta; l_i - l_j)/2}\right)}{\sqrt{\widehat{\text{Var}}(\theta^{k_1, k_2}(l_i, l_j))}} \right)^2.$$

## 5 The post-processing procedure

As the bivariate Brown-Resnick process model developed in this paper describes the joint distribution of the observed and forecasted maxima of the variable of interest, it allows for some spatial post-processing of the original forecast. In this section, we will describe the resulting post-processing procedure in more detail and provide some tools to verify the procedure and the underlying model.

### 5.1 Post-processing via conditional simulation

Let  $\hat{\xi}^{\text{obs}}, \hat{\mu}_v^{\text{obs}}(\cdot), \hat{\sigma}_v^{\text{obs}}(\cdot), \hat{\xi}^{\text{pred}}, \hat{\mu}_v^{\text{pred}}(\cdot), \hat{\sigma}_v^{\text{pred}}(\cdot)$  and  $\hat{\vartheta}$  be estimates for the GEV and variogram parameters derived from past training data. Further, assume that we have  $v_{\max}^{\text{pred}}(l_i, p), \hat{m}(l_i, p)$  and  $\hat{s}(l_i, p), i = 1, \dots, n_l$ , based on forecasts for  $n_l$  locations  $l_1, \dots, l_{n_l}$  and a time period  $p$  in near future. Then, we obtain an arbitrary number  $K$  of realizations  $(v_j(l_i))_{1 \leq i \leq n_l}, j = 1, \dots, K$ , of the modeled distribution of the maximal observation conditional on the forecast by the following three-step procedure:

1. Transform  $v_{\max}^{\text{pred}}(\cdot, p)$  to standard Gumbel margins:

$$x^{\text{pred}}(\cdot) = \frac{1}{\hat{\xi}^{\text{pred}}} \log \left( 1 + \hat{\xi}^{\text{pred}} \frac{v_{\max}^{\text{pred}}(\cdot, p) - \hat{\mu}_v^{\text{pred}}(\cdot, p)}{\hat{\sigma}_v^{\text{pred}}(\cdot, p)} \right),$$

where  $\hat{\mu}_v^{\text{pred}}$  and  $\hat{\sigma}_v^{\text{pred}}$  are given by Eq. 13 for  $k = \text{pred}$ .

2. Conditional simulation of a bivariate Brown-Resnick process given its second component: Simulate  $K$  independent realizations  $(x_j^{\text{obs}}(\cdot), x_j^{\text{pred}}(\cdot)), j = 1, \dots, K$ , of a bivariate Brown-Resnick process associated to the pseudo cross-variogram  $\gamma(\hat{\vartheta}^{\text{obs}}; \cdot)$  with standard Gumbel margins conditional on  $x_j^{\text{pred}}(\cdot) = x_j^{\text{pred}}(\cdot)$ .
3. Transform  $x_j^{\text{obs}}(\cdot)$  to GEV margins: For  $j = 1, \dots, K$ , set

$$v_j(\cdot, p) = \hat{\sigma}_v^{\text{obs}}(\cdot, p) \frac{\exp(\hat{\xi}^{\text{obs}} x_j^{\text{obs}}(\cdot)) - 1}{\hat{\xi}^{\text{obs}}} + \hat{\mu}_v^{\text{obs}}(\cdot, p),$$

where  $\hat{\mu}_v^{\text{obs}}$  and  $\hat{\sigma}_v^{\text{obs}}$  are given by Eq. 13 for  $k = \text{pred}$ .

The random fields obtained by this three-step procedure can be interpreted as post-processed probabilistic forecasts for the maxima of the variable of interest at time period  $p$ . While the first and the third steps only consist of marginal transformations, the conditional simulation in the second step is the challenging part of the procedure. For this step, the algorithm by Dombry et al. (2013) can be used. Note that the algorithm, which has originally been designed for conditional simulation of univariate Brown-Resnick processes, can directly be transferred to the multivariate case by perceiving the multivariate processes as univariate processes on a larger index set. However, the computations will be computationally expensive, in particular if the number of conditioning locations gets large.

## 5.2 Verification

In practical applications, the proposed post-processing procedure and the underlying model need to be verified. Here, we do not only consider the full bivariate Brown-Resnick model which forms the base of the post-processing procedure, but also intermediate models such as the marginal GEV model and the univariate model. This allows us to evaluate the effect of incorporating the spatial dependence structure and the forecasted maxima, respectively.

For the evaluation and verification of the different models, we choose a standard verification score in probabilistic prediction, the (negatively oriented) continuous ranked probability score (CRPS) (cf. Gneiting and Raftery 2007):

$$CRPS(F, x) = \int_{-\infty}^{\infty} |y - x| F(dy) - \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y_1 - y_2| F(dy_1) F(dy_2),$$

where  $F$  is a real-valued distribution and  $x \in \mathbb{R}^m$  is an observation. The continuous ranked probability score is a strictly proper scoring rule, i.e.  $\int CRPS(F, x) F(dx) \leq \int CRPS(G, x) F(dx)$  for all distribution functions  $F$  and  $G$  with finite first moments and equality if and only if  $F = G$ . This indicates that the mean CRPS for different observations is the smaller, the better the predicted distribution  $F$  fits to the true distribution of the observation data. The usefulness of the CRPS for evaluating extremes was shown by Friederichs and Thorarinsdottir (2012).

First, we evaluate the improvement in predictive quality by fitting the parameters of the GEV model given in Eqs. 12 and 13 to the observations instead of the forecast, i.e. we calculate  $CRPS^{\text{obs}}(l_i)$  and  $CRPS^{\text{pred}}(l_i)$  where

$$CRPS^k(l_i) = n_d^{-1} \sum_{d=1}^{n_d} CRPS(G_{\hat{\xi}_i^k, \hat{\mu}_i^k(l_i, d), \hat{\sigma}_i^k(l_i, d)}, v_{\max}^{\text{obs}}(l_i, d))$$

for every station  $l_i$ ,  $1 \leq i \leq n_l$ , and  $k \in \{\text{"obs"}, \text{"pred"}\}$ . For the calculation, we employ the closed formula for the CRPS of a GEV provided by Friederichs and Thorarinsdottir (2012). For  $\xi \neq 0$ , they obtain

$$\begin{aligned} CRPS(G_{\xi, \mu, \sigma}, x) &= \left( x - \mu + \frac{\sigma}{\xi} \right) (2F(x) - 1) \\ &\quad - \frac{\sigma}{\xi} (2^\xi \Gamma(1 - \xi) - 2\Gamma_l(1 - \xi, -\log F(x))) \end{aligned}$$

where  $\Gamma_l$  is the lower incomplete gamma function. Furthermore, the CRPS for the GEV fitted to the observations can be compared to the CRPS of the original forecast

$$CRPS^{\text{orig}}(l_i) = n_p^{-1} \sum_{p=1}^{n_p} CRPS(F_{l_i, p}^{\text{orig}}, v_{\max}^{\text{obs}}(l_i, p)), \quad (19)$$

where  $F_{l_i, p}^{\text{orig}}$  denotes the distribution of the original (probabilistic) forecast for the maximum of the variable of interest at location  $l_i$  within time period  $p$ . If this forecast is given by an ensemble of values, such as the output of a numerical weather prediction model, for example,  $F_{l_i, p}^{\text{orig}}$  corresponds to the empirical distribution function of this sample. If the forecast corresponds to a single value,  $CRPS^{\text{orig}}(l_i)$  reduces to the mean absolute error.

Finally, the full bivariate model and, thus, the proposed post-processing procedure can be evaluated and verified by considering the CRPS

$$\text{CRPS}^{\text{biv}}(l_i) = n_p^{-1} \sum_{p=1}^{n_p} \text{CRPS}(F_{l_i, p | v_{\max}^{\text{pred}}}, v_{\max}^{\text{obs}}(l_i, p))$$

where  $F_{l_i, p | v_{\max}^{\text{pred}}}$  denotes the distribution of the observed maximum at location  $l_i$ ,  $1 \leq i \leq n_l$ , within time period  $p$  conditional on  $v_{\max}^{\text{pred}}$ , i.e. the distribution of the post-processed forecast, with the CRPS of the original forecast,  $\text{CRPS}^{\text{orig}}(l_i)$ .

## 6 Application to real data

In this section, we will apply the fitting and verification procedure described in Section 4 to real wind gust data consisting both of observation and forecast data. We will see that, even though the marginal distributions are fitted quite well, a forecast based on the single GEV for the observations is not able to outperform the forecast by the numerical weather prediction model. However, the results for the bivariate model indicate that the post-processing procedure proposed in Section 5.1 improves the predictive quality. We also discuss the uncertainty of the obtained estimates.

### 6.1 The data

We consider observed as well as forecasted wind speed data provided by Germany's National Meteorological Service, the Deutscher Wetterdienst (DWD). We use observations from 218 DWD weather stations over Germany at 360 days from March 2011 to February 2012. The weather stations register mean and maximum wind speed on an hourly basis. Due to the inertia of the measuring instruments, the maximum wind speed approximately corresponds to the highest 3-second average wind speed. Here, we use the maximum wind speed  $v_{\max}^{\text{obs}}(l, d)$  between 08 UTC and 18 UTC for each station  $l$  and each day  $d$ .

Furthermore, for each day, forecasts for the wind speed maxima and for the hourly mean wind speed both in 10m height above ground and for the 10-hour-period from 08 UTC to 18 UTC are available. The forecasts are provided by the COSMO-DE ensemble prediction system (EPS) operated by DWD. COSMO-DE (Baldauf et al. 2011) is a non-hydrostatic limited-area numerical weather prediction model that gives forecasts for the next 21 hours on a horizontal grid with a width of 2.8km covering Germany and neighboring countries. For each variable of interest, the COSMO-DE EPS yields forecasts consisting of 20 ensemble members stemming from COSMO-DE runs with five different physical parameterizations and four different lateral boundary conditions provided by global model forecasts. For more details on the Consortium for Small-scale Modeling see <http://www.cosmo-model.org/>, and Gebhardt et al. (2011) and Peralta et al. (2012) for COSMO-DE EPS.

The COSMO-DE EPS is initialized every 3 hours. Here, we take the forecasts that are initialized at 00 UTC. Using the forecasts for the nearest grid point of a

station, we obtain forecasts  $v_{\text{mean}}^{(1)}(l, d, \tau), \dots, v_{\text{mean}}^{(20)}(l, d, \tau)$ ,  $\tau \in \{9, 10, \dots, 18\}$ , and  $v_{\text{max}}^{(1)}(l, d), \dots, v_{\text{max}}^{(20)}(l, d)$  for every weather station  $l$  and every day  $d$ . Here,  $v_{\text{mean}}^{(j)}(l, d, \tau)$  and  $v_{\text{max}}^{(j)}(l, d)$  denote the forecast for the mean wind speed between  $(\tau - 1)$  UTC and  $\tau$  UTC and the maximal wind speed, respectively, at station  $l$  and day  $d$ , forecasted by the  $j$ th COSMO-DE ensemble member.

For the application of our model with a stationary spatial dependence structure, in the following, we will restrict ourselves to forecasted and observed data for 119 DWD stations north of  $51^\circ\text{N}$ , denoted by  $l_1, \dots, l_{119}$ , as the northern part of Germany has a much more homogeneous topography than the southern part.

## 6.2 Applying the univariate model

As the wind speed observations correspond to 3-second averages, the daily maximal wind gusts  $v_{\text{max}}^{\text{obs}}$  can be perceived as the maximum of a long time series. Further, the distribution of a single wind speed is frequently modeled by a Weibull or a Gamma distribution (e.g., Conradsen et al 1984; Pavia and O'Brien 1986; Slougher et al 2007), that is, given a fixed shape parameter of the Weibull or Gamma distribution, respectively, which is spatially and temporally constant, the single observations may be assumed to come from a location-scale family of distributions. These considerations give support to the usage of the GEV model presented in Section 2 as a model for the maximal wind speed  $V_{\text{max}}^k(l_i, d)$ ,  $i \in \{1, \dots, 119\}$ ,  $d \in \{1, \dots, 360\}$ . Fitting a GEV distribution to the standardized wind speeds  $y^k(l_i, d)$  as defined in Eq. 10 needs the estimates  $\hat{m}(l_i, d)$  and  $\hat{s}(l_i, d)$  for the mean and the standard deviation of the underlying wind speed distribution. We aim to extract these characteristics from the forecast. Here, instead of direct estimates for the mean and the standard deviation, we use

$$\hat{m}(l_i, d) = \frac{20}{\max_{j=1}^{20}} \frac{1}{10} \sum_{\tau=9}^{18} v_{\text{mean}}^{(j)}(l_i, d, \tau) \quad (20)$$

$$\text{and } \hat{s}(l_i, d) = \left( \frac{1}{199} \sum_{j=1}^{20} \sum_{\tau=9}^{18} \left[ v_{\text{mean}}^{(j)}(l_i, d, \tau) - \frac{1}{200} \sum_{j'=1}^{20} \sum_{\tau'=9}^{18} v_{\text{mean}}^{(j')}(l_i, d, \tau') \right]^2 \right)^{1/2}. \quad (21)$$

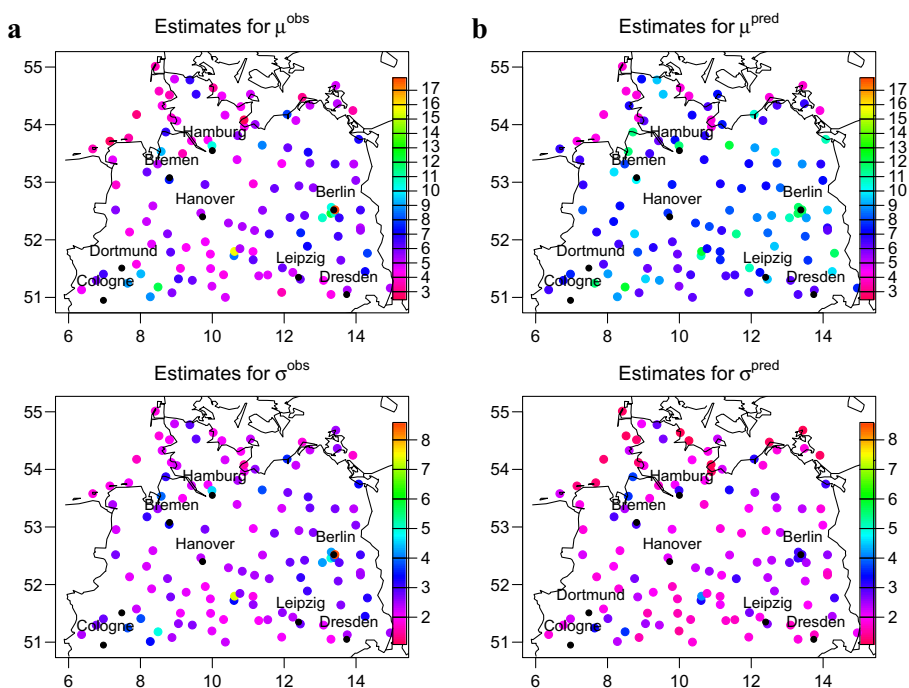
Even though not providing consistent estimates for mean and standard deviation of the underlying distribution,  $\hat{m}(l_i, d)$  and  $\hat{s}(l_i, d)$  lead to a consistent normalization in the following sense: If the data  $v_{\text{max}}^{\text{obs}}$  and the forecasts  $v_{\text{mean}}^{(j)}$  are affinely transformed (i.e. the parameters  $m$  and  $s$  are modified) in the same way, the normalized data  $y^{\text{obs}}(l_i, d)$  remain unchanged. This choice of  $\hat{m}(l_i, d)$  and  $\hat{s}(l_i, d)$  also ensures the identifiability of the GEV parameters  $\mu^k(l_i)$  and  $\sigma^k(l_i)$ .

As described in Section 4, the GEV parameters for the standardized observations can be estimated via maximum likelihood and the hypotheses that these are spatially constant can be checked via Kolmogorov-Smirnov tests. For  $\xi^{\text{obs}}$ , we obtain a  $p$ -value of 0.194. The analogous tests for  $\mu^{\text{obs}}$  and  $\sigma^{\text{obs}}$  both yield  $p$ -values smaller than  $2.2 \cdot 10^{-16}$ . Thus, the hypotheses that the residuals of the estimates of  $\mu^{\text{obs}}$  and  $\sigma^{\text{obs}}$

follow a normal distribution both can be rejected and, consequently, we stick to the assumption that the GEV parameters, location and scale, differ among the stations.

In contrast to the location and the scale parameter, the shape parameter of the GEV will be assumed to be spatially constant in northern Germany with the value  $\xi^{\text{obs}} = \hat{\xi}^{\text{obs}} = \frac{1}{119} \sum_{i=1}^{119} \hat{\xi}^{\text{obs}}(l_i) = 0.043$  (empirical standard deviation of the sample  $\{\hat{\xi}^{\text{obs}}(l_i) : i = 1, \dots, 119\}$ : 0.049). Note, however, that the estimated shape parameter differs significantly (to a 5%-level) from the mean value in case of 20 stations. For six of these stations, it even differs highly significantly (to a 1%-level), and four of them even to a 0.1%-level. The parameter estimates  $\hat{\mu}(l_i)$  and  $\hat{\sigma}(l_i)$ ,  $1 \leq i \leq 119$  for the location and scale parameters, respectively, obtained by maximum likelihood estimation with fixed shape parameter  $\xi^{\text{obs}} = \hat{\xi}^{\text{obs}}$  are depicted in Fig. 1a. Note that the estimated vectors of location and scale parameters show a strong empirical correlation of 0.97. By Eq. 3, the data can be transformed to standard Gumbel margins. Kolmogorov-Smirnov tests performed separately for each station yield  $p$ -values of at least 0.098 with a mean value of 0.718 which indicates that the GEV model fits quite well for all the stations.

As a fit of the GEV distribution to the forecast is needed for both verification of the marginal model and the bivariate Brown-Resnick model, we repeat our analysis



**Fig. 1** **a** Estimates  $\hat{\mu}^{\text{obs}}(l_i)$  and  $\hat{\sigma}^{\text{obs}}(l_i)$  for the location and scale parameters corresponding to the observed maximal wind speed at the stations in the northern part of Germany. **b** Estimates  $\hat{\mu}^{\text{pred}}(l_i)$  and  $\hat{\sigma}^{\text{pred}}(l_i)$  for the location and scale parameters corresponding to the forecasted maximal wind speed at the stations in the northern part of Germany

replacing the observed maximal wind speed  $v_{\max}^{\text{obs}}(l_i, d)$  by  $v_{\max}^{\text{pred}}(l_i, d)$ , i.e. a forecast for the maximal wind speed at station  $l_i$  and day  $d$ . Here, we use the maximum over the 20 corresponding COSMO-DE ensemble members

$$v_{\max}^{\text{pred}}(l_i, d) = \max_{j=1, \dots, 20} v_{\max}^{(j)}(l_i, d), \quad 1 \leq i \leq 119, \quad 1 \leq d \leq 360,$$

which suggests that the distribution of  $v_{\max}^{\text{pred}}$  should be close to a GEV distribution. Note that this choice of  $v_{\max}^{\text{pred}}$  is in complete accordance to the choice of  $\hat{m}(l_i, d)$  as maximal mean of all the ensemble members in Eq. 20.

As the Kolmogorov-Smirnov test of the normalized estimates for  $\xi^{\text{pred}}$  yields a  $p$ -value of 0.53 and the estimates differ significantly from the mean for seven stations (for three of them very significantly), we may assume a shape parameter of  $\xi^{\text{pred}} = \hat{\xi}^{\text{pred}} = \frac{1}{119} \sum_{i=1}^{119} \xi^{\text{pred}}(l_i) = 0.028$  (empirical standard deviation of the sample  $\{\hat{\xi}^{\text{pred}}(l_i) : i = 1, \dots, 119\}$ : 0.044) at every station in northern Germany. However, the hypotheses that the estimates for the location and the scale parameter follow a normal distribution have been both rejected. The maximum likelihood estimates  $\hat{\mu}^{\text{pred}}(l_i)$  and  $\hat{\sigma}^{\text{pred}}(l_i)$ ,  $1 \leq i \leq 119$ , with fixed shape parameter are shown in Fig. 1b. Here, the empirical correlation of the vectors of estimated location and scale parameters is just as strong as in case of the observations. Kolmogorov-Smirnov tests of the transformed data  $x^{\text{pred}}(l_i, d)$  for every station yield  $p$ -values of at least 0.142 with and equal 0.748 in average which also indicates an appropriate fit.

The spatial dependence is modeled by a univariate Brown-Resnick process which is obtained by a weighted least squares fit of the extremal coefficient function. Here, the weights depend on the variance of the estimators  $\hat{\nu}^{\text{obs}}(l_i, l_j)$  (see Section 4) estimated by a jackknife procedure where the extremal coefficients are reestimated leaving out one month of data. The estimated extremal coefficients  $\hat{\theta}^{\text{obs}}$  and the fitted extremal coefficient function

$$\tilde{\theta}^{\text{obs}}(s, t) = 2\Phi\left(\sqrt{\gamma_{\hat{\theta}^{\text{obs}}}(s-t)/2}\right), \quad s, t \in \mathbb{R}^2.$$

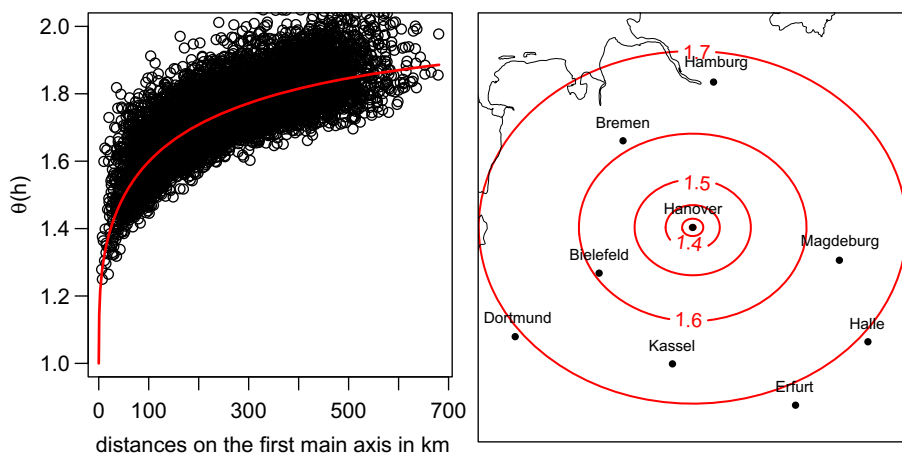
are displayed in Fig. 2. Here, the estimated coefficients seem to be fitted quite well.

For verification, we first calculate the mean CRPS for each of the two models given by Eq. 12,  $\text{CRPS}^{\text{obs}}(l_i)$  and  $\text{CRPS}^{\text{pred}}(l_i)$ , for every station  $l_i$ ,  $1 \leq i \leq 119$ . Then, the improvement or deterioration by using the GEV distributions of the observations instead of the forecasts is expressed in terms of the skill score (e.g., Gneiting and Raftery 2007)

$$S(l_i) = 1 - \frac{\text{CRPS}^{\text{obs}}(l_i)}{\text{CRPS}^{\text{pred}}(l_i)}$$

which has the value 1 in case of an “optimal” model which equals  $v_{\max}^{\text{obs}}$  a.s. and the value 0 if both models yield the same result. Here,  $S(l_i) > 0$  for 115 of 119 stations.





**Fig. 2** *Left:* The estimated extremal coefficients  $\hat{\theta}^{\text{obs}}$  (black circles) and the fitted extremal coefficient function  $\hat{\theta}^{\text{obs}}$  (red line) of the normalized random field  $X^{\text{obs}}(\cdot, d)$  of observed wind gusts. *Right:* Contour level plot of the fitted extremal coefficient function  $\hat{\theta}^{\text{obs}}(l_0, \cdot)$  where  $l_0$  is located at Hanover

For the skill score corresponding to the mean CRPS averaged over all the stations, we obtain

$$S = 1 - \frac{\sum_{i=1}^{119} \text{CRPS}_{l_i}^{\text{obs}}}{\sum_{i=1}^{119} \text{CRPS}_{l_i}^{\text{pred}}} \approx 0.293.$$

Note that, for simplicity, the reference model (12) for the predictions is based on the maximal ensemble members  $v_{\max}^{\text{pred}}(l_i, d)$  only and further information given by the maximal wind speed forecasted by the other ensemble members are neglected. Thus, we further compare the CRPS of the GEV model for the observations,  $\text{CRPS}^{\text{obs}}(l_i)$ , with the CRPS of the original COSMO-DE ensemble,  $\text{CRPS}^{\text{orig}}(l_i)$ , taking the ensemble forecast as a probabilistic forecast with equal probability for each ensemble member. Here, the skill  $\tilde{S}(l_i) = 1 - \text{CRPS}^{\text{obs}}(l_i)/\text{CRPS}^{\text{orig}}(l_i)$  is positive for 37 of 119 only, with the skill of the averaged CRPS being approximately  $-0.032$ . As the skill score is slightly negative, the COSMO-DE ensemble forecast seems to contain more information than our marginal model.

Note that, for a fair comparison, we should avoid the validation of our model on the same data that have been used for the model fit. Hence, we perform cross validation: Separately for every month, the GEV parameters are reestimated leaving out the data for this month and using only the data for the other eleven months for the model fit. The GEV parameters estimated for different months in this way show very little variation corroborating the assumption that they are constant in time. Further, the verification results above are confirmed: We obtain skill scores of 0.285 for the CRPS compared with the GEV model for the forecast and  $-0.048$  compared to the COSMO-DE ensemble.

### 6.3 Applying the bivariate model

A bivariate Brown-Resnick process is fitted to the transformed data according to Section 4. Here, as a preliminary analysis suggests, the parameter  $\rho$  is set to the maximal value yielding a valid variogram, i.e.

$$\rho = \frac{(1 + v_{12}^{-1})^{v_{12}+1}}{(1 + v_1^{-1})^{\frac{1}{2}v_1+\frac{1}{2}}(1 + v_2^{-1})^{\frac{1}{2}v_2+\frac{1}{2}}}.$$

The estimate  $\hat{\vartheta}$  of the remaining eleven pseudo cross-variogram parameters leads to the fitted extremal coefficient function

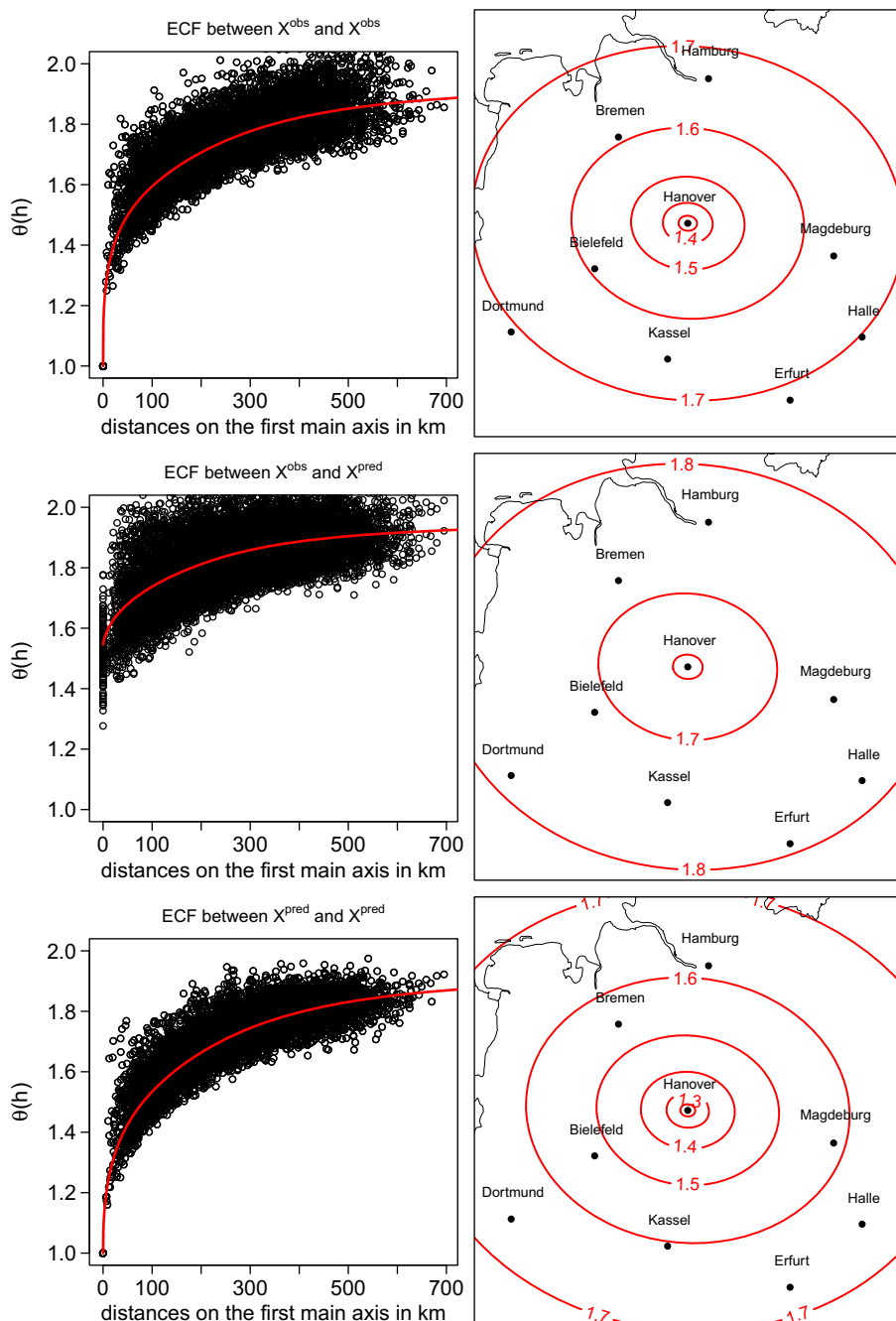
$$\begin{aligned}\tilde{\theta}(l_i, l_j) &= \left( \tilde{\theta}^{k_1, k_2}(l_i, l_j) \right)_{k_1, k_2 \in \{\text{"obs"}, \text{"pred"}\}} \\ &= 2 \left( \Phi \left( \sqrt{\gamma_{k_1 k_2}(\hat{\vartheta}; l_i - l_j) / 2} \right) \right)_{k_1, k_2 \in \{\text{"obs"}, \text{"pred"}\}}.\end{aligned}$$

Figure 3 presents the estimated extremal coefficients  $\hat{\theta}^{k_1, k_2}(l_i, l_j)$ , and the fitted extremal coefficient functions  $\tilde{\theta}^{k_1, k_2}(\cdot, \cdot)$  for  $k_1, k_2 \in \{\text{"obs"}, \text{"pred"}\}$ . As illustrated, the fitted model seems to be appropriate with respect to the behavior of the extremal coefficient function. Figure 4 depicts a simulated realization of the corresponding Brown-Resnick process associated to the variogram  $\gamma(\hat{\vartheta}; \cdot)$  with standard Gumbel margins. The realization indicates a remarkable amount of positive correlation between  $x^{\text{obs}}$  and  $x^{\text{pred}}$  which emphasizes the gain of information by taking  $x^{\text{pred}}$  into account.

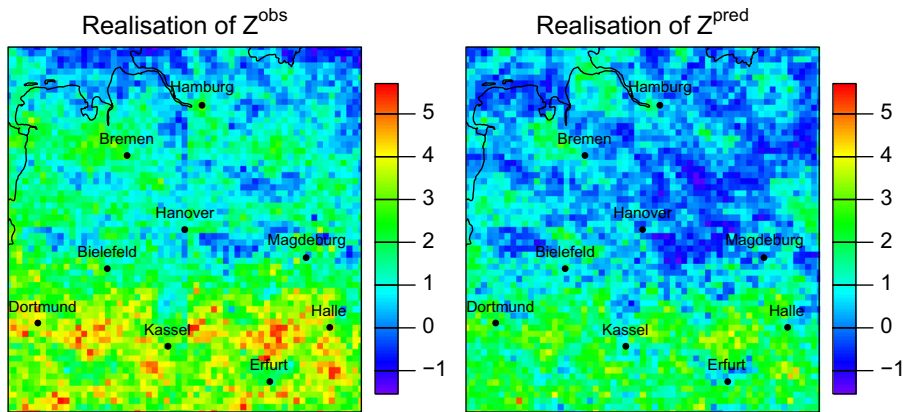
In order to verify the bivariate model, we apply the post-processing procedure proposed in Section 5.1. Due to the computational complexity of the conditional simulation, we do not simulate the observations at all stations simultaneously conditional on the forecast at all locations, but perform post-processing with sample size  $K = 20$ , i.e. the size of the original COSMO-DE ensemble, at each location separately conditioning on the forecast at the same location and two neighboring grid cells only. We calculate the CRPS of the post-processed distribution,  $\text{CRPS}^{\text{biv}}(l_i)$ , and compare it with  $\text{CRPS}^{\text{orig}}(l_i)$ , i.e. the CRPS belonging to the empirical distribution of the COSMO-DE ensemble, yielding a positive skill score for 82 of 119 stations where the skill score related to the mean CRPS equals 0.128 (0.111 cross-validated). If we increase the sample size  $K = 100$ , we obtain an improved skill score of 0.164 (0.147 cross-validated), being positive for 104 of 119 stations. Thus, we conclude that the post-processing procedure based on the bivariate Brown-Resnick model is able to improve the forecast given by COSMO-DE ensemble.

### 6.4 Uncertainty assessment

We assess the uncertainty in the estimation of the model parameters via parametric bootstrap, i.e. we simulate data sets from the fitted model and repeat the estimation procedure. As a detailed analysis of the numerical model producing the forecasts is beyond the scope of our paper, we do not account for the uncertainty in the estimates  $\hat{m}(l_i, d)$  and  $\hat{s}(l_i, d)$ , but focus on the normalized values  $\{(v_{\max}^{\text{obs}}(l_i, d) -$



**Fig. 3** Left: The estimated extremal coefficients (black circles) and the fitted extremal coefficient function (red line) of the normalized bivariate random field  $(X^{\text{obs}}, X^{\text{pred}})$  of observed and forecasted wind gusts. Right: Contour level plots of the fitted extremal coefficient function  $\tilde{\theta}(l_0, \cdot)$  where  $l_0$  is located at Hanover



**Fig. 4** Simulated realization of a Brown-Resnick process associated to the variogram  $\gamma(\vartheta; \cdot)$  with standard Gumbel margins

$\hat{m}(l_i, d)/\hat{s}(l_i, d), (v_{\max}^{\text{obs}}(l_i, d) - \hat{m}(l_i, d))/\hat{s}(l_i, d) : i = 1, \dots, 119, d = 1, \dots, 360\}$ . To this end, we draw 360 independent realizations from the bivariate Brown-Resnick process fitted in Sections 6.2 and 6.3 using the simulation algorithm by Dombry et al. (2016). The procedure is repeated 100 times yielding 100 independent data sets of the same size as the original one. Following the steps described in Sections 6.2 and 6.3, we thus obtain 100 independent estimates.

The sample of estimates  $\hat{\xi}^{\text{obs}}$  has mean 0.043 and a standard deviation of 0.02. In order to validate the  $p$ -value of the Kolmogorov-Smirnov tests for  $\xi^{\text{obs}}$ , we repeat these tests on the simulated data sets and obtain a  $p$ -value smaller than the original one (0.194) in 22 of 100 cases which supports the non-rejection of the hypothesis that  $\xi^{\text{obs}}$  is spatially constant. The  $p$ -values of the tests for  $\mu^{\text{obs}}$  and  $\sigma^{\text{obs}}$  are confirmed, as well.

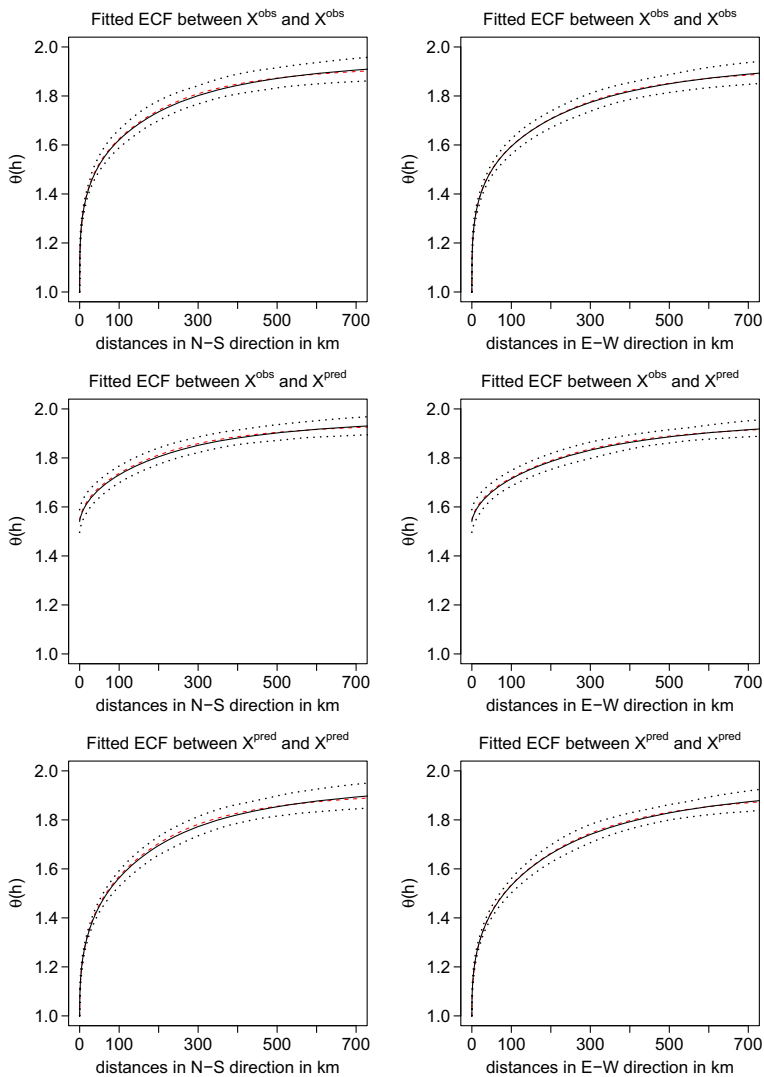
Analogously, the results for the marginal parameters for the forecast are verified: The sample of estimates  $\hat{\xi}^{\text{pred}}$  has mean 0.03 and a standard deviation of 0.02, while the original  $p$ -value (0.53) of the test for  $\xi^{\text{obs}}$  is undercut in 62 of 100 cases.

Further, we assess the estimation of the Brown-Resnick model parameters. The original parameter values and the sample means and standard deviations obtained from the parametric bootstrap are presented in Table 1. It can be seen that most parameters are recovered well by the estimation procedure. For some parameters of

**Table 1** Results for parameters of the bivariate Brown-Resnick model obtained via parametric bootstrap

	$b$	$\theta$	$\beta$	$\sigma$	$\kappa$	$c$	$\sigma_1$	$\nu_1$	$\sigma_2$	$\nu_2$	$a$
orig.	1.2	-0.10	0.91	2.0	425	1.5	1.3	0.3	1.1	0.4	124
mean	1.2	0.05	0.72	1.3	307	1.5	1.4	0.3	1.2	0.5	306
std. dev.	0.2	0.36	0.23	0.8	214	0.1	0.4	0.5	0.4	0.7	283

the bivariate Whittle-Matérn model, however, rather large variances are observed, a phenomenon which is often encountered in practice in accordance to the fact that not all parameters of the Whittle-Matérn model can be consistently estimated from observations in a fixed domain (cf. Zhang 2004). Even though some parameters show considerable variation, there is little variation in the bivariate extremal coefficient function as displayed in Fig. 5.



**Fig. 5** Results for the bivariate extremal coefficient functions obtained via parametric bootstrap as functions of the distance in north-south (*left*) and east-west (*right*) direction. The dashed lines are the true extremal coefficient functions while the solid lines gives the pointwise bootstrap sample mean of the fitted extremal coefficients. The dotted lines indicate the pointwise 0.025 and 0.975-quantiles, respectively

Besides the uncertainty of the parameter estimates, we also assess the deviation of the non-parametrically estimated extremal coefficients from the parametrically estimated extremal coefficient function. To this end, for  $k_1, k_2 \in \{\text{"obs"}, \text{"pred"}\}$ , we calculate the root-mean-square error

$$\text{RMSE}^{k_1, k_2} = \left[ \frac{1}{119^2} \sum_{1 \leq i, j \leq 119} \left( \hat{\theta}^{k_1, k_2}(l_i, l_j) - 2\Phi(\sqrt{\gamma_{k_1, k_2}(\hat{\vartheta}; l_i, l_j)/2}) \right)^2 \right]^{1/2}.$$

For the original data, we obtain the root-mean-square errors  $\text{RMSE}^{\text{obs, obs}} = 0.074$ ,  $\text{RMSE}^{\text{obs, pred}} = \text{RMSE}^{\text{pred, obs}} = 0.078$  and  $\text{RMSE}^{\text{pred, pred}} = 0.061$ . Repeating the calculation for the simulated data sets, we obtain sample mean root-mean-square errors 0.039 (for  $\text{RMSE}^{\text{obs, obs}}$ ), 0.041 (for  $\text{RMSE}^{\text{obs, pred}}$  and  $\text{RMSE}^{\text{pred, obs}}$ ) and 0.037 (for  $\text{RMSE}^{\text{pred, pred}}$ ) with standard deviations 0.0018, 0.0019 and 0.0018, respectively. Thus, the deviations for the real data are roughly twice as large as expected in case of the model being correct. One may conclude, that the extremal coefficient for real data is not only a function of the distance, but may be modulated by the topography or the climate at the stations.

Finally, note that the same methodology could be used to assess the uncertainty of the post-processed forecast. To this end, different sets of estimated parameters could be used as input for the post-processing procedure described in Section 5.1 to determine the uncertainty of its output. However, a full analysis would also require an assessment of the uncertainty in the estimates  $\hat{m}(l_i, d)$  and  $\hat{s}(l_i, d)$  which, as already mentioned above, is beyond the scope of this work.

**Acknowledgments** This research has been financially supported by VolkswagenStiftung within the project “Mesoscale Weather Extremes – Theory, Spatial Modeling and Prediction (WEX-MOP)”. The research of M. Oesting has also partially been funded by the ANR project ‘McSim’. Observational data and COSMO-DE-EPS forecasts have been kindly provided by Deutscher Wetterdienst in Offenbach, Germany. The authors are grateful to the Associate Editor and an anonymous referee for their valuable suggestions improving this article.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., Reinhardt, T.: Operational convective-scale numerical weather prediction with the COSMO model. *Mon. Wea. Rev.* **139**, 3887–3905 (2011)
- Berg, C., Christensen, J.P.R., Ressel, P.: *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. Springer, New York (1984)
- Brasseur, O.: Development and application of a physical approach to estimating wind gusts. *Mon. Wea. Rev.* **129**(1), 5–25 (2001)
- Brown, B.M., Resnick, S.I.: Extreme values of independent stochastic processes. *J. Appl. Probab.* **14**(4), 732–739 (1977)
- Chilès, J.P., Delfiner, P.: *Geostatistics Modeling Spatial Uncertainty*, 2nd edn. Wiley, Hoboken (2012)

- Clark, I., Basinger, K.L., Harper, W.V.: MUCK—a novel approach to co-kriging. In: Geostatistical, sensitivity, and uncertainty methods for ground-water flow and radionuclide transport modeling. Battelle Press, San Francisco (1989)
- Coles, S.: Regional modelling of extreme storms via max-stable processes. *J. R. Statist. Soc. Ser. B* **55**, 797–816 (1993)
- Coles, S.: An Introduction to Statistical Modeling of Extreme Values. Springer, London (2001)
- Coles, S., Tawn, J.: Modelling extremes of the areal rainfall process. *J. R. Statist. Soc. Ser. B* **58**, 329–347 (1996)
- Conradsen, K., Nielsen, L.B., Prahm, L.P.: Review of Weibull statistics for estimation of wind speed distributions. *J. Clim. Appl. Meteorol.* **23**, 1173–1183 (1984)
- Cooley, D., Naveau, P., Poncet, P.: Variograms for Spatial Max-Stable Random Fields. In: Dependence in Probability and Statistics, pp. 373–390. Springer, New York (2006)
- Dombry, C., Éyi-Minko, F., Ribatet, M.: Conditional simulation of max-stable processes. *Biometrika* **100**(1), 111–124 (2013)
- Dombry, C., Engelke, S., Oesting, M.: Exact simulation of max-stable processes. *Biometrika* **103**(2), 303–317 (2016)
- Engelke, S., Malinowski, A., Kabluchko, Z., Schlather, M.: Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Statist. Soc. Ser. B* **77**(1), 239–265 (2015)
- Friederichs, P., Thorarindottir, T.L.: Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics* **23**(7), 579–594 (2012)
- Gebhardt, C., Theis, S., Paulat, M., Ben Bouallègue, Z.: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.* **100**(2), 168–177 (2011)
- Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M.: Handbook of Spatial Statistics. CRC Press, Boca Raton (2010)
- Genton, M.G., Padoan, S.A., Sang, H.: Multivariate max-stable spatial processes. *Biometrika* **102**(1), 215–230 (2015)
- Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102**(477), 359–378 (2007)
- Gneiting, T., Kleiber, W., Schlather, M.: Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105**(491), 1167–1177 (2010)
- Guttorp, P., Gneiting, T.: Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika* **93**(4), 989–995 (2006)
- Handcock, M.S., Wallis, J.R.: An approach to statistical spatial-temporal modeling of meteorological fields. *J. Amer. Statist. Assoc.* **89**(426), 368–378 (1994)
- Huser, R., Davison, A.C.: Space-time modelling of extreme events. *J. R. Statist. Soc. Ser. B* **76**(2), 439–461 (2014)
- Kabluchko, Z.: Extremes of independent Gaussian processes. *Extremes* **14**(3), 285–310 (2011)
- Kabluchko, Z., Schlather, M., de Haan, L.: Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37**(5), 2042–2065 (2009)
- Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and Related Properties of Random Sequences and Processes. Springer, New York (1983)
- Molchanov, I., Stucki, K.: Stationarity of multivariate particle systems. *Stochastic. Process. Appl.* **123**(6), 2272–2285 (2013)
- Papritz, A., Künsch, H., Webster, R.: On the pseudo cross-variogram. *Math. Geol.* **25**(8), 1015–1026 (1993)
- Pavia, E.G., O’Brien, J.J.: Weibull statistics of wind speed over the ocean. *J. Clim. Appl. Meteorol.* **25**, 1324–1332 (1986)
- Peralta, C., Ben Bouallègue, Z., Theis, S.E., Gebhardt, C., Buchhold, M.: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.* **117**, D07,108 (2012)
- Ribatet, M.: Spatial extremes: Max-stable processes at work. *J. Soc. Fr. Stat.* **154**(2), 156–177 (2013)
- Schilling, R.L., Song, R., Vondraček, Z.: Bernstein Functions: Theory and Applications. Gruyter, Berlin (2010)
- Schlather, M., Tawn, J.A.: A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90**(1), 139–156 (2003)

- Slougher, J.M.L., Raftery, A.E., Gneiting, T., Fraley, C.: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.* **135**, 3209–3220 (2007)
- Smith, R.L.: Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**(1), 67–90 (1985)
- Smith, R.L.: Max-stable processes and spatial extremes, unpublished manuscript (1990)
- Stein, M.L.: *Interpolation of Spatial Data: Some Theory for Kriging* Springer Series in Statistics. Springer, New York (1999)
- Zhang, H.: Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99**(465), 250–261 (2004)